



Summer temperature extremes in Europe: how does the definition affect the results?

Agnieszka Sulikowska^{1,2} · Agnieszka Wypych^{2,3}

Received: 10 May 2019 / Accepted: 24 February 2020 / Published online: 19 March 2020
© The Author(s) 2020

Abstract

Nearly every study dealing with temperature extremes underscores the lack of a universal and broadly used method of identifying such events. The most popular are relative methods, which are based on the empirical distribution of temperature at each location (i.e., percentiles). The aim of this study was to evaluate the effects of the various percentile-based methods of defining hot days on the analysis of their frequency of occurrence, trends, and geographic patterns in summer in Europe. The basis for the research consists of daily maximum (TX) and minimum (TN) values of air temperature for 1961–2017 for Europe obtained from the E-OBS database. A hot day occurs when air temperature exceeds the 90th percentile-based threshold. These thresholds are determined using the following: (I) various temperature metrics (TX and TN), (II) various baseline periods (1961–1990, 1971–2000, 1981–2010), and (III) different timeframes within the year that the percentile is calculated for (summer season, separate summer months, and each calendar day). Our results indicate that the use of different variants of the percentile-based definition leads to differences in the geographic patterns of frequencies of and trends in summer hot days in Europe. The differences are especially substantial within the results obtained using various temperature metrics and baseline periods, and they are relatively small when different timeframes within the year that the percentile is calculated for are considered. On the example of the case study, we also show how the use of different research approaches may affect the intensity and spatial extent of an extreme temperature event.

Keywords Heat extremes · Percentile threshold · Hot days · Spatial distribution · Climate change

1 Introduction

Extreme temperature events, hereinafter termed temperature extremes, or days with extremely high or low air temperature and their sequences, may have serious societal, agricultural, economic, and ecological impacts. Long-lasting heat waves are considered to be one of the natural hazards that are most dangerous to human health and life; their effect on the increase in morbidity and mortality has been shown in various regions

of the world (Xu et al. 2016). Research on the variability and consequences of temperature extremes is very popular due to observed changes in their frequency of occurrence (Fu-Min et al. 2018). The lack of a generally recognized and used definition, and what it entails, a method for the identification of such events, is emphasized in almost every climatological publication that delves into these issues (Perkins 2015). The IPCC (2012) defines “a climate extreme (extreme weather or climate event)” as “the occurrence of a value of a weather or climate variable above (or below) a threshold value near the upper (or lower) end of the range of observed values of the variable.” In accordance with this definition, it is assumed in climatological research that air temperature is “extreme” when it reaches or is higher (lower) than an assumed threshold value. Temperature extremes that are so defined are amenable to various types of statistical analysis, and it is possible to determine their various attributes, such as their frequency of occurrence, spatial range, and duration (Stephenson 2008).

The threshold, above/below which air temperature is considered to be extreme, may be determined in many ways,

✉ Agnieszka Sulikowska
a.sulikowska@doctoral.uj.edu.pl

¹ Department of Ecology, Climatology and Air Protection, University of Agriculture in Kraków, Al. Mickiewicza 24/28, 30-059 Kraków, Poland

² Department of Climatology, Jagiellonian University, 7 Gronostajowa St, 30-387 Kraków, Poland

³ Institute of Meteorology and Water Management, National Research Institute, 61 Podleśna St, 01-673 Warszawa, Poland

and the choice of the method depends on the research goal. The simplest solution is to choose a constant absolute threshold, which may be related to impacts. However, temperature thresholds so determined are relevant in particular geographic regions and in specified time periods only; their values change with latitude and climate characteristics, and also depending on the season of the year (Stephenson 2008; IPCC 2012). This is the reason why relative thresholds became popular, especially those based on the empirical distribution of temperature at each studied location, that is, on percentiles. This approach ensures that a given part of air temperature observations (for example, 10% if using the 90th percentile) is “extreme” by definition. The advantage of this method is the possibility of comparing results obtained in geographic areas with a different climate and also in different seasons of the year. Limitations include the fact that the frequency of occurrence is assumed to be known and that extremes that are so identified are not necessarily “extreme” because of their impact (Zhang et al. 2011; Ustrnul et al. 2012; IPCC 2012). This method is recommended by the IPCC (2012) and WMO (2009).

A very large number of percentile-based indices can be found in the literature. These indices are based on various percentiles, temperature metrics, and baseline periods; moreover, percentiles are calculated for different timeframes within the year, which makes comparing results difficult. Temperature extremes are most often determined using percentiles ranging from the 90th to the 99th (for example, Alexander et al. 2006; Moberg et al. 2006; Fischer and Schär 2010; Lhotka and Kysely 2015a; Hoy et al. 2017), although the use of other percentiles, such as the 75th (for example, Carril et al. 2008) or the 80th (for example, Della-Marta et al. 2007b) may also be found. The choice of a rigorous criterion, for example, the 99th percentile, increases the probability of the identification of events of presumed serious consequence to society and the environment. On the other hand, an analysis of changes in the frequency of occurrence of such events carries a considerable amount of uncertainty due to their rarity (Zhang et al. 2011). Choosing a relatively lenient criterion, for example, the 90th percentile, ensures selecting a sufficiently large sample for analysis of changes with time; however, it may come to light that relatively many “non-extreme” events will be included in the research sample. Therefore, the choice of percentile is often a compromise between having a sufficient number of cases and the level of how extreme the cases really are (Perkins and Alexander 2013).

Warm temperature extremes are identified using either maximum (TX), minimum (TN), or diurnal average (TG) air temperature. Research in which relationships between warm extremes and morbidity and mortality are considered often takes into account indices based on TN because it is the events during which TN remains high and does not allow humans to recuperate that are the most severe for some populations (Xu

et al. 2016). In strictly climatological analyses, warm extremes are increasingly often determined based on both TX and TN, sometimes classifying them as “daytime” or “nighttime” events, respectively (e.g., Busuioc et al. 2015; Efthymiadis et al. 2011; Kažys et al. 2011; Spinoni et al. 2015). The physical basis for this discrimination is that diurnal extreme temperatures are not necessarily related to the same physical processes as extremes in nocturnal cooling. And so, for example, Spinoni et al. (2015) distinguished “warm days,” based on TX, from “warm nights,” based on TN, whereas Lavaysse et al. (2018) identified “hot days” using TX and TN separately and then finding their intersection (both TX and TN exceeded the 90th percentile). Still another approach was presented by Revich and Shaposhnikov (2008) who defined “extremely hot days” using TG. On the other hand, in research where both warm as well as cold extremes are considered, it is frequently so that TX serves to determine the warm, whereas TN cold extremes (e.g., Lhotka and Kysely 2015b; Sulikowska et al. 2018).

Researchers use various baseline periods to determine thresholds, above or below which air temperature is considered extreme. Baseline periods that are most frequently used are as follows: 1961–1990 (e.g., Alexander et al. 2006; Rusticucci et al. 2016; Hoy et al. 2017; Wypych et al. 2017), 1971–2000 (e.g., Spinoni et al. 2015; Sanderson et al. 2017), and 1981–2010 (e.g., Russo et al. 2015). Other, less standard periods are also used, and they often comprise the entire considered time period (e.g., Stefanon et al. 2012; Lavaysse et al. 2018; Tomczyk et al. 2018). According to the latest recommendations of the WMO, the baseline period should be “the most recent 30-year period terminating in a year ending with a zero” (1981–2010 at the time of writing). Only in the case of research areas focused on climate variability assessment and climate change monitoring, which include research on the temporal variability of temperature extremes, should the stable baseline period of 1961–1990 be used (WMO 2017).

In some cases, percentile-based indices used by various research teams may seem the same or similar because they are based on the same percentile, temperature metric, and baseline period. However, when one takes a closer look, it comes to light that the percentiles were calculated for different timeframes within the year. And so, for example, Tomczyk and Bednorz (2016) and Sanderson et al. (2017) calculated a percentile for an entire calendar year, Lhotka and Kysely (2015a) determined a percentile for the summer season (JJA), and Revich and Shaposhnikov (2008) did it separately for each summer month. Yet, another approach is to determine a percentile for each calendar day separately using an x -day-window-centered method. This method allows taking into consideration the variability of air temperature during the year. For example, Efthymiadis et al. (2011) used a 5-day window, Della-Marta et al. (2007a) used a 15-day window, and Stefanon et al. (2012) used a 21-day window.

Seemingly, the same criterion (extreme TX = TX > 95th percentile) was used in every one of the cited studies, but the top 5% of diurnal maxima (TX's) were identified out of samples of different size and value range.

Attempts to compare different research approaches pertain especially to one, special kind of temperature extreme, which is heat waves. Smith et al. (2013) and later You et al. (2017) have shown there are substantial differences in the frequency, trends, and geographic patterns of heat waves in the USA and China, respectively, which were defined with the use of 16 indices. A similar approach was used by Fenner et al. (2018) who identified substantial differences in trends in heat waves determined using 10 different definitions in 125-year-long data series from Berlin and Potsdam. Perkins and Alexander (2013) focused on a comparison of trends in various characteristics of heat waves determined with the use of the 90th percentile of TX and TN in Australia and found that generally, they are quite similar in both sign and spatial extent.

Nevertheless, there have not been any publications thus far, which would be focused on the comparison of the definitions of temperature extremes, which would be more basic on the one hand, yet more multifaceted on the other hand. In this context, “basic” would mean considering days with an extremely high air temperature, with no regard to the fact whether they form sequences (i.e., heat waves, warm spells). The term “multifaceted” is to be understood as an attempt to evaluate how the analysis and results are influenced by a change in the various components of a percentile-based definition.

Therefore, the aim of this study was to evaluate the effects of the various methods of determination of percentile-based thresholds on the frequency of occurrence, trends, and geographic patterns of summer hot days in Europe. More specific research goals include an evaluation of the role of the use of various temperature metrics as well as different baseline periods and timeframes within the year that the percentile was calculated from.

2 Data and methods

The study is based on diurnal maximum (TX) and minimum (TN) air temperatures in Europe in the years 1961–2017 obtained from the E-OBS gridded dataset with a spatial resolution of $0.5^\circ \times 0.5^\circ$ (version 17; Haylock et al. 2008; www.ecad.eu). In this dataset, TX and TN are defined as the 24h maximum and minimum, respectively (Lavaysse et al. 2018). In the analyses, only grid points without missing data were used. Most analyses were conducted using TX, the availability of which is shown in Fig. 1. To highlight local characteristics, particular attention was paid to the grid point with the coordinates 20.25° E, 50.25° N, which is located in Central Europe, in the vicinity of the city of Kraków, Poland (herein after referred to as “KRK”; Fig. 1).

A hot day is defined as a day on which air temperature exceeds the 90th percentile of the local probability density function (IPCC 2012). Analyses were performed for summer (JJA) hot days in the period 1961–2017. Three different components of a hot day definition were examined: temperature metrics, baseline periods, and timeframes. The focus was kept on an evaluation of the influence of different approaches on the frequency of occurrence and trends of hot days with particular attention paid to spatial patterns. Trends were determined using the nonparametric Mann-Kendall test and their statistical significance was assessed using Sen's slope estimator at the significance level $\alpha = 0.05$ (von Storch and Zwiers 2003).

In the first part of the study, the role of using different temperature metrics was evaluated. Hot days were identified using TX and TN along with their intersection when both TX and TN exceed the 90th percentile. The percentile was calculated for the period 1961–1990 using a 15-day window centered on each calendar day (explained below; Table 1). TG was not used, as information on temperature conditions for a 24-h period that it provides was too general, and it is rarely used in analyses of daily temperature extremes.

In the second part, the effects of different baseline periods were examined. Hot days were identified using the 90th percentile of TX, calculated for WMO (2017) recommended and most often used 30-year-long periods: 1961–1990, 1971–2000, and 1981–2010 (Table 1).

In the third part of the study, different timeframes for percentile calculation were examined. The 90th percentile was calculated using TX from the following: (I) the whole summer (JJA), (II) separate summer months, and for each calendar day using (III) a 15- and (IV) a 5-day window-centered method. In the case of a percentile calculation for each calendar day using a 5-day window method, a probability density function was computed for day X using temperature data for 30-year climatology between $X - 2$ days and $X + 2$ days. One proceeds similarly in the case of a time window of different sizes (Fischer and Schär 2010; Stefanon et al. 2012; Perkins and Alexander 2013). The number of days the percentile was calculated from was different in each of these approaches: 2760 days in the case of a seasonal percentile, 900/930 days in the case of monthly percentiles, and 450 and 150 in the case of two variants of daily percentiles. The annual 90th percentile, that is calculated using TX values from an entire calendar year, was not included in the analysis, as it was close to the median of TX for summer, and consequently reflected average, instead of extremely warm, temperature conditions (Zhang et al. 2011). Higher annual percentiles were successfully used to identify hot days by Tomczyk and Bednorz (2016) and Sanderson et al. (2017), among others.

Intensity and spatial extent are undoubtedly some of the most important attributes of hot days, which affect their severity and outcomes (Stephenson 2008; Horton et al. 2016). In

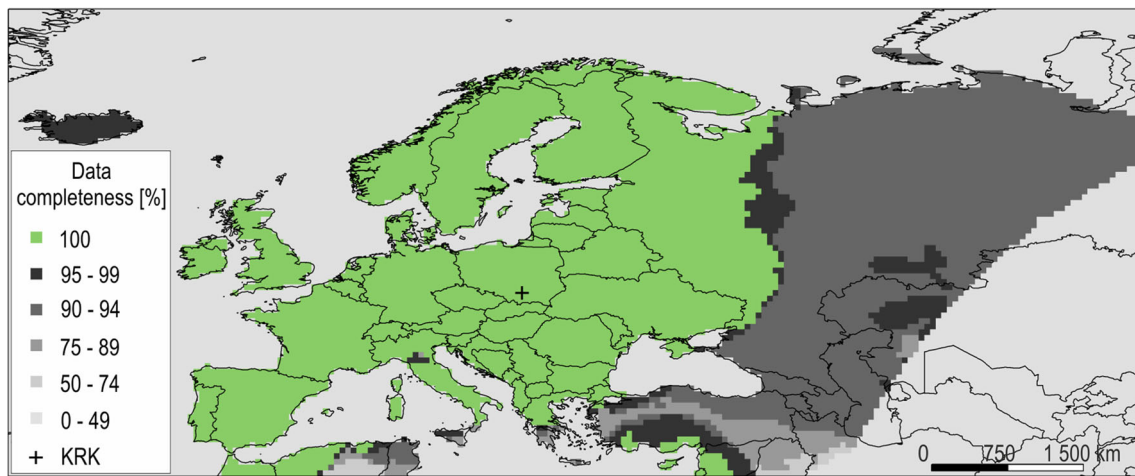


Fig. 1 Availability of diurnal TX values in the summer in Europe in the E-OBS gridded dataset, version 17, for the period 1961–2017

the last part of the study, it was evaluated how the use of the considered indices makes the spatial range and intensity of a hot day different. This was accomplished on the example of a heat event on August 31, 2015, in Central Europe (Hoy et al. 2017; Wypych et al. 2017). The spatial range (total area (TA)) was defined by the number of grid points where an extreme temperature occurred, whereas intensity was characterized using the cumulative temperature excess above the percentile-based threshold (total intensity (TI)):

$$TI = \sum_{i=1}^N (TX - TX_{90})$$

where TX is the maximum daily air temperature, TX_{90} is the corresponding 90th percentile, and N is the number of grid points where $TX > TX_{90}$ (based on Kysely 2010 and Wypych et al. 2017). A similarly developed formula was used for TN. HD_{INT} was not considered because it uses both TX and TN, which made it impossible to calculate the total intensity according to the formula above.

3 Results

3.1 Hot day occurrence

3.1.1 Temperature metrics

To determine hot extremes, the daily maximum (HD_{TX}) as well as minimum (HD_{TN}) air temperatures may be used. The average numbers of HD_{TN} and HD_{TX} in the considered area are comparable, 14 and 12 hot days, respectively, but their spatial distributions differ substantially (Fig. 2a, b). HD_{TN} is characterized by large spatial differentiation and by relatively high values over southern Europe (up to 30 hot days per summer), while the geographical distribution of HD_{TX} is much less variable. The differences between the number of HD_{TN} and HD_{TX} may be positive or negative, with the highest discrepancy in the Mediterranean region (Fig. 3). When using both temperature metrics, that is, when a hot day is defined as a day when both TX and TN exceed the 90th percentile (HD_{INT}), the number of events is smaller by 50% and 55%

Table 1 Definitions of indices used to identify hot days, which were used in this study (details in the text)

| Research phase (section in the paper) | Definition component and its variants | | Index name | Remaining components of definition |
|--|---------------------------------------|------------------------|------------|---|
| 1 (Section 3.1.1.) | Temperature metrics | TX | HD_{TX} | 90th percentiles calculated for the period 1961–1990 using a 15-day window centered on every calendar day |
| | | TN | HD_{TN} | |
| | | TX and TN intersection | HD_{INT} | |
| 2 (Section 3.1.2.) | Baseline periods | 1961–1990 | HD_{61} | 90th percentiles of TX, calculated using a 15-day window centered on every calendar day |
| | | 1971–2000 | HD_{71} | |
| | | 1981–2010 | HD_{81} | |
| 3 (Section 3.1.3.) | Timeframes | Seasonal | HD_S | 90th percentiles of TX, calculated for the period 1961–1990 |
| | | Monthly | HD_M | |
| | | Daily: 15-day window | HD_{D5} | |
| | | Daily: 5-day window | HD_{D15} | |

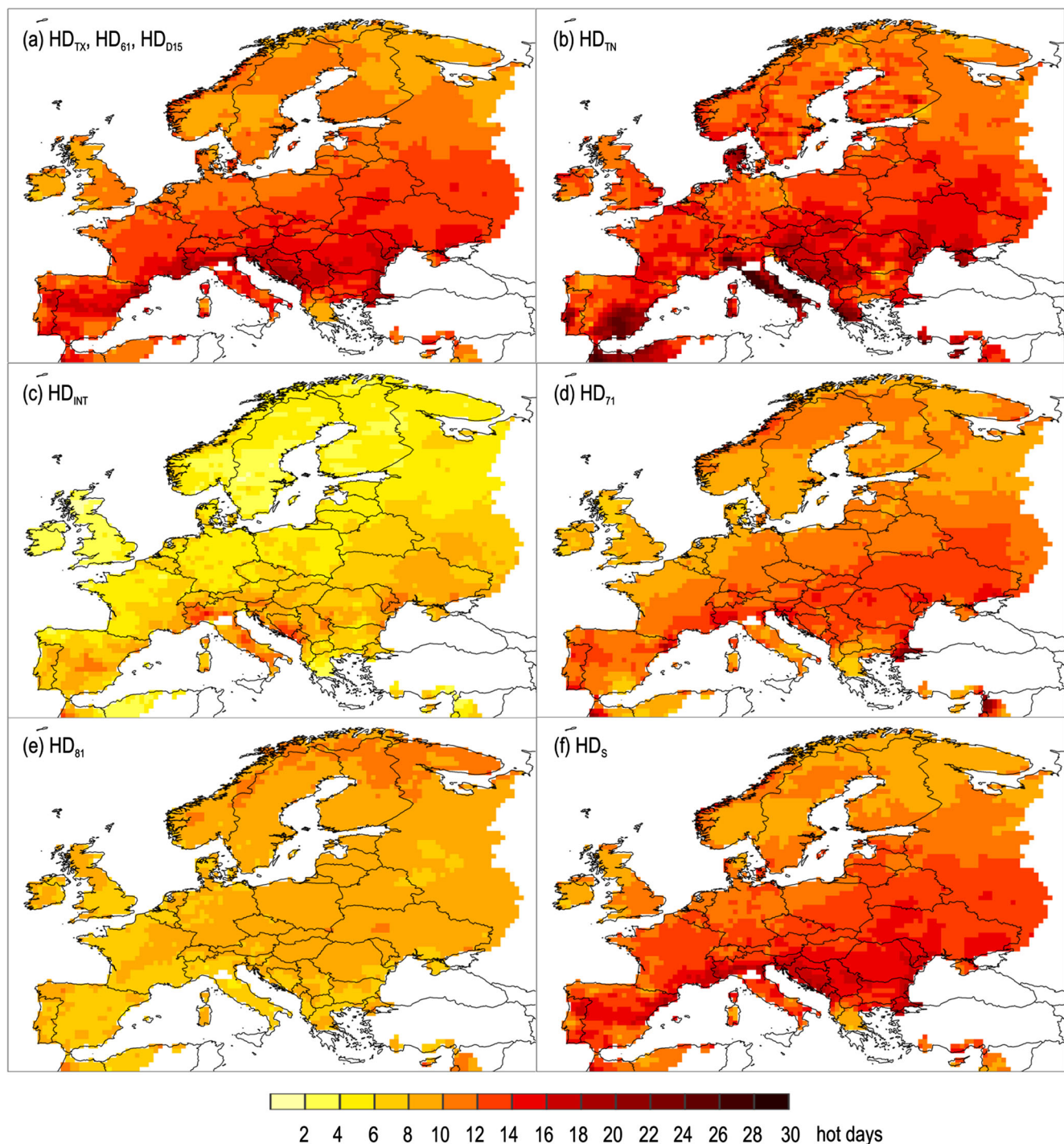


Fig. 2 The average number of hot days in the summer during the period 1961–2017 identified using the indices: **a** HD_{TX} , HD_{61} , and HD_{D15} , **b** HD_{TN} , **c** HD_{INT} , **d** HD_{71} , **e** HD_{81} , and **f** HD_S

than HD_{TX} and HD_{TN} , respectively (Fig. 2c). This means that about half of threshold exceedances by TX (TN) are associated with threshold exceedances by TN (TX).

3.1.2 Baseline periods

It may be expected that temperature thresholds determined for each subsequent baseline period will be higher due to climate

warming. However, this is not always the case. Daily thresholds determined at KRK for the 1971–2000 period are nearly the same as those for the oldest baseline period for 14% of summer days, or even lower, for 17% of summer days (Fig. 4). This is due to the high frequency of cold summers in Poland in the years 1971–2000 (Wypych et al. 2017). As seen in Fig. 4, the daily pattern of the 90th percentile changes significantly from one baseline period to another, both in terms of the

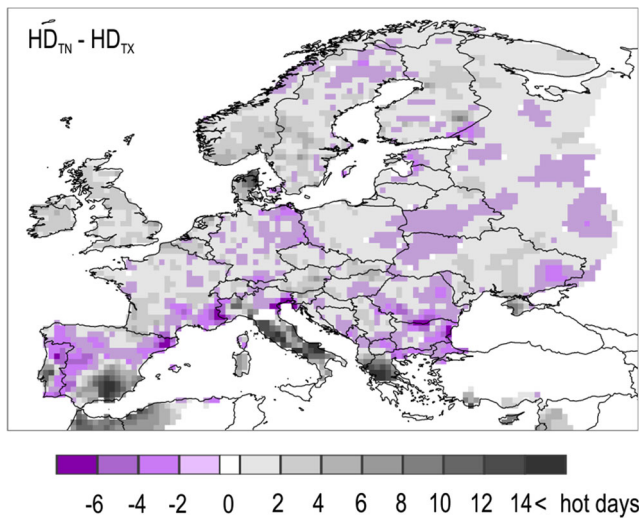


Fig. 3 Differences in the average number of hot days in the summer during the period 1961–2017 identified using the indices HD_{TX} and HD_{TN}

magnitude of threshold values as well as the moment of occurrence of their maxima and minima.

The number of hot days calculated for the baseline period 1961–1990 (HD_{61}) and that calculated for 1971–2000 (HD_{71}) is greatest for Southern Europe, whereas hot days determined using the baseline period 1981–2010 (HD_{81}) occur most frequently in northern regions (Fig. 2a, d, e). The average number of hot days ranges from 5 to 8 at a minimum and from 13 to 21 at a maximum, depending whether HD_{61} or HD_{71} or HD_{81} is considered. As expected, due to climate warming, in most of Europe, the average number of hot days is greater when an older baseline period, and thus lower threshold, is used. However, it is the other way around in some areas, primarily in the North (Fig. 2a, d, e). These are the effects of regional TX variations during baseline periods, as shown in the KRK example above. Sometimes, as in the Kola Peninsula

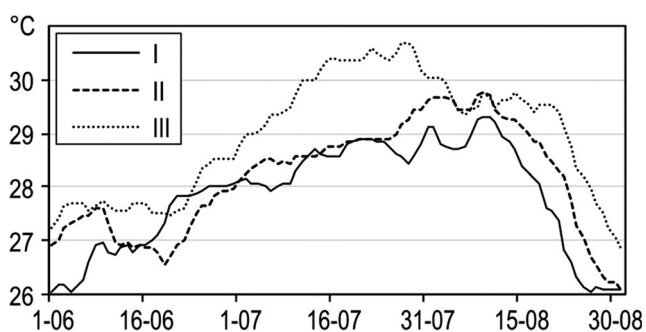


Fig. 4 Time series of the 90th percentile of TX in summer determined using a 15-day moving window centered on every calendar day at the grid point KRK for three periods: (I) 1961–1990, (II) 1971–2000, (III) 1981–2010

in Russia, the older the baseline period, the higher the percentile-based threshold. Consequently, in this region, the average number of hot days increases when a younger baseline period is used (Fig. 2a, d, e).

It follows from the assumptions of the method that in the case of using a given percentile based on a given baseline period, the frequency of events is the same regardless of considered geographic area. Meanwhile, differences within Europe are substantial (Fig. 2a, d, e). The spatial variation of the frequency of hot days is a consequence of the spatial distribution of temperature during both the baseline period and beyond it. Temperature conditions during the baseline period determine spatial variation of thresholds necessary to be exceeded to qualify a given day as a hot day. On the other hand, temperature conditions before and after the baseline period bring high or low numbers of threshold exceedances in each given geographic region. The spatial variation of hot days is also affected by the occurrence of air temperatures that are exactly equal to the threshold value, as noted in Zhang et al. (2011) and elsewhere.

3.1.3 Timeframes

The choice of timeframe for percentile calculation is determined by the events one is interested in. In order to illustrate the properties of the four indices considered here, the total number of hot days determined on each summer day at KRK along with 90th percentile-based thresholds is shown in Fig. 5. A constant value of the seasonal percentile (Fig. 5a) causes that the greatest number of HD_S is found in the warmest time of the summer. This works similarly in the case of monthly percentiles, but the greatest number of events occurs in the warmest part of each month (Fig. 5b). Only daily percentiles reflect air temperature variability within a season and owing to this, the number of hot days during summer keeps fluctuating around some value (Fig. 5c, d). The principal difference between them is the magnitude of day-to-day variability, which is larger in the case of the threshold based on a 5-day window than a 15-day one (Fig. 5c, d).

The differences described above are not observable in the spatial distributions of hot days, which reveal very similar geographical patterns for all indices in this group. For this reason, only results for hot days distinguished using the seasonal percentile (HD_S) and for those based on percentiles calculated for each calendar day using a 15-day window-centered method (HD_{D15}) are presented in Fig. 2. On average, 12 hot days occur every summer over the entire studied area, and their number fluctuates from 7–8 to 20–21, depending on the index used (Fig. 2a, f). Generally, the differences are largest between HD_S and other indices; however, they rarely exceed 1 day and reach 3 days at most.

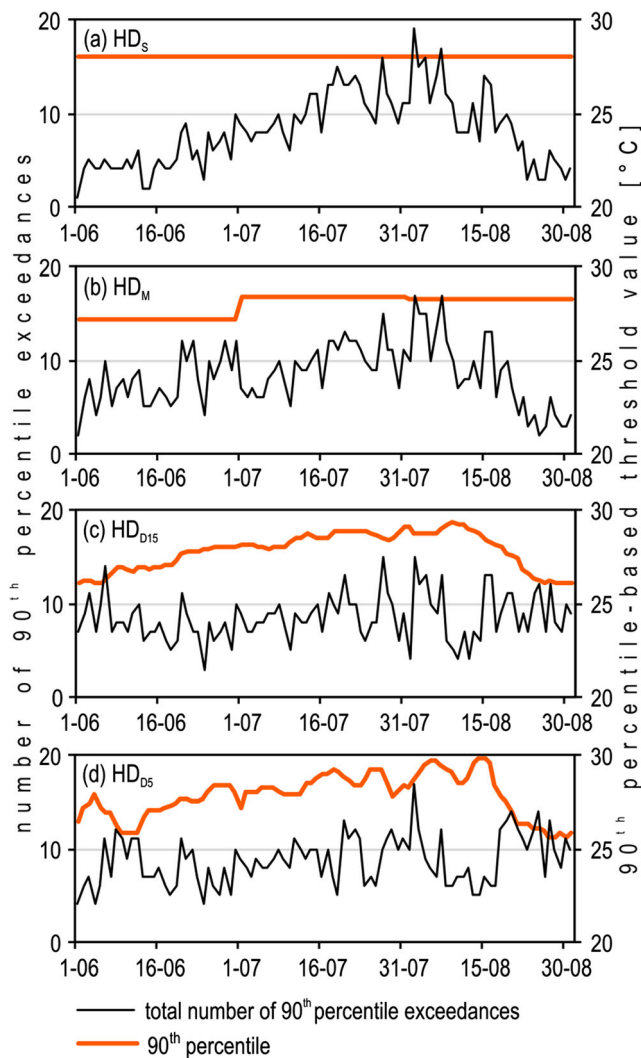


Fig. 5 Total number of 90th percentile exceedances calculated for each summer day in the period 1961–2017 depending on the index used: **a** HD_s , **b** HD_m , **c** HD_{D15} , and **d** HD_{D5} at the grid point KRK

3.2 Long-term trends

Differences between the examined indices are readily observable in temporal variability analysis. The number of hot days identified using TN and TX and their intersection is characterized by growing trends in a greater part of the considered area; however, their rates of change and spatial distribution are significantly different (Fig. 6a, b, c). HD_{TN} is characterized by the highest and HD_{INT} by the lowest maximum growth rates (14 and 5 hot days per 10 years, respectively). Spatially, the rate of change of HD_{TN} varies the most, while the rate of change of HD_{INT} varies the least. In the case of HD_{TN} , 99% of trends are positive, and 92% are increases that are statistically significant (Fig. 6b). Growth trends in HD_{TX} and HD_{INT} account for 86 to 87%, but in the case of the latter, the share of statistically significant increases is larger (Fig. 6a, c).

Differences in trends of HD_{TN} and HD_{TX} do vary geographically reaching 14 hot days per 10 years, with regions of higher rates of change, for either HD_{TN} or HD_{TX} . The increase rate of HD_{INT} is typically slower than those of HD_{TX} or HD_{TN} . A decreasing trend of HD_{INT} occurred at one grid point only. In the case of HD_{TN} , there were nine such isolated points, whereas HD_{TX} decreased in several small areas in Europe (2.4% of grid points), although the changes are mostly not significant. Therefore, trends in these geographic areas are of opposite sign—depending on which index is used (Fig. 6a, b, c).

In the case of indices based on percentiles calculated over different baseline periods, the spatial distribution of trends is generally similar (Fig. 6a, d, e). Along with the use of later baseline periods, the share of grid points decreases, in which the number of hot days increases and vice versa. In the case of positive trends, generally, the later the baseline period, the slower the rate of change. This results directly from climate warming and the fact, that air temperature distribution is close to normal. The same rate of warming results in a larger increase in the number of hot days identified using lower thresholds (i.e., older baseline periods), which are closer to the median of the temperature distribution, than those determined using higher thresholds (i.e., more recent baseline periods), which are closer to the end of the distribution tail. The area where no changes are observed increases along with the use of later baseline period (Fig. 6a, d, e). No grid points were found, at which trends have a different direction depending on the index used.

Spatial distributions of trends in the number of hot days based on percentiles computed over different timeframes are very similar (Fig. 6a, f). The share of positive trends comprises 85 to 87%, while statistically significant trends include 59 to 61% of all grid points. Decreasing trends comprise 2 to 3% and in individual cases they are statistically significant. Differences in the growth rate of the number of hot days between the considered indices mostly do not exceed 1 hot day per 10 years. There are only a few grid points which are characterized by different directions of change depending on which index is being used.

3.3 Extreme temperature event on August 31, 2015

To assess the influence of different indices on those attributes of hot days which primarily affect their impacts, a case study focusing on total intensity (TI; °C) and total area (TA; number of grid points with threshold exceedance) of a hot day is presented. These attributes are considered for all the indices examined in this paper (except HD_{INT} , see Section 2).

On August 31, 2015, the maximum air temperature (TX) in Europe was highest in its central part and exceeded 34 °C in

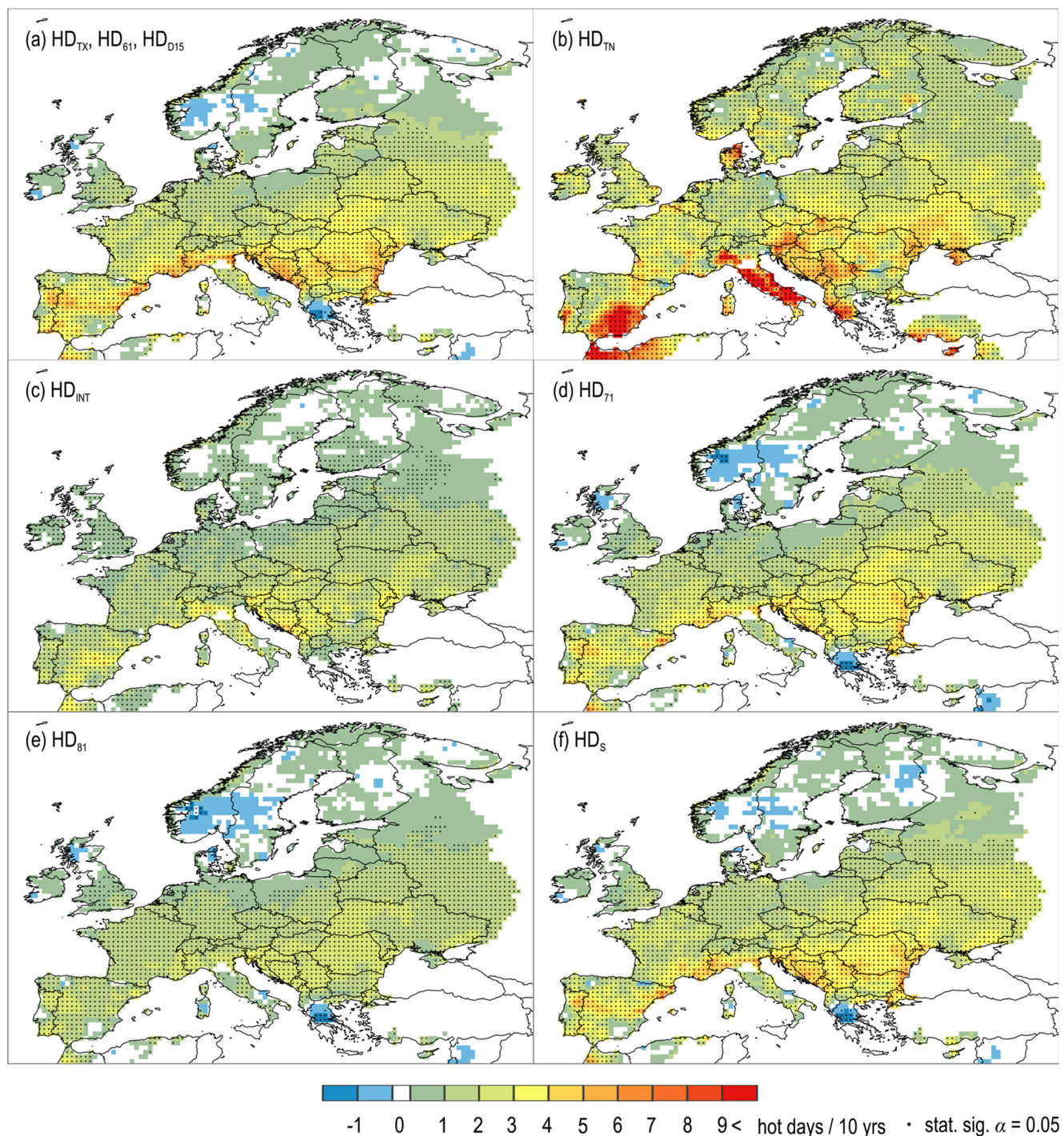


Fig. 6 Trends of the number of hot days in summer in the period 1961–2017 (hot days per 10 years) determined using the indices: **a** HD_{TX} , HD_{61} , and HD_{D15} , **b** HD_{TN} , **c** HD_{INT} , **d** HD_{71} , **e** HD_{81} , and **f** HD_S

the area extending from the eastern border of Germany to Western Ukraine and from Central Poland to Northern Bulgaria, excluding the mountain areas (Fig. 7a). It was hottest in the Balkans, where TX was over 37 °C. Daily minimum air temperature (TN) in Europe was highest in the southern part of the continent, exceeding 20 °C in Spain and Italy, and

also in the Balkans and on the eastern coast of the Black Sea (Fig. 7b). TN was mostly 16–18 °C in the central part of the continent.

Most considered methods, and actually all based on TX, indicate that the extreme temperature event occurred primarily in Central Europe (Fig. 8). In the case of indices based on

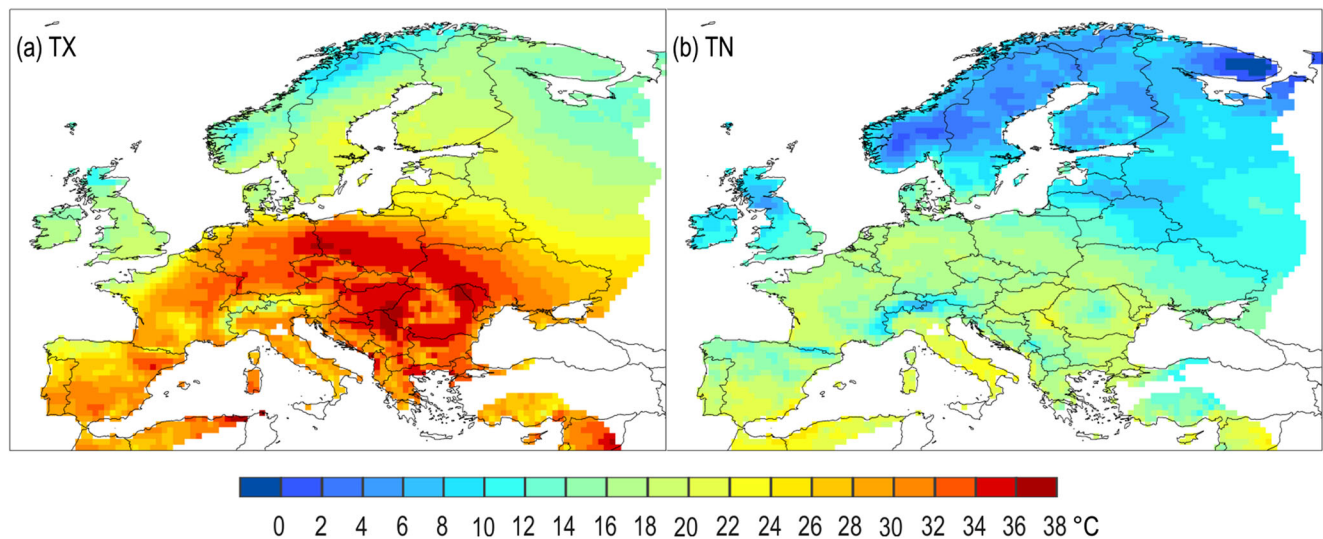


Fig. 7 Diurnal **a** maximum and **b** minimum air temperatures in Europe on August 31, 2015

different temperature metrics, the differences primarily pertain to the area of the occurrence of extremely high temperature and TI. When HD_{TX} is used, the area affected by the extreme temperature is compact, i.e., the entire region of Central Europe experienced a high temperature excess which decreased towards the edges. As minimum air temperature is more sensitive to local conditions, the spatial extent of the extreme temperature event spreads out more. TI, which in total is over 40% lower in the case of HD_{TN} in comparison with HD_{TX} , is spatially variable (Fig. 8b). A comparison of indices based on different baseline periods confirms previous observations—differences between HD_{61} and HD_{71} are small (Fig. 8a, c), whereas TI and TA for HD_{81} are lower, being 82–87% of what was obtained with the use of older baseline periods (Fig. 8d). Among indices based on percentiles calculated over different timeframes, the results for HD_S are slightly higher than those for HD_M (Fig. 8e, f). TI and TA values obtained for both HD_S and HD_M are 59–74% of what was obtained for indices based on daily percentiles (Fig. 8a, e, f, g). The differences between the latter are not substantial, but TI and TA are higher in the case of HD_{D5} . Of course, it should be emphasized that the relationships between the obtained results are not constant and they may look completely different if an extreme event occurring in another part of the season is considered.

4 Discussion and conclusions

It has been demonstrated in this paper that the use of different determination methods affects the analysis of the frequency, trends, and variability of extreme temperature events. This was accomplished on the example of hot days in summer in Europe in the years 1961–2017. Hot days are defined as days on which air temperature exceeds the 90th percentile of the

local probability density function. Three groups of indices were established based on several variants of the hot day definition, i.e., by employing various temperature metrics, baseline periods, and timeframes for a percentile calculation. The comparison of results obtained within these groups leads to the conclusion that a change in any component of a definition of a hot day has an effect on climatological analyses; however, the importance of these effects varies depending on the component modified. Nevertheless, all indices reveal similar climate signals, which are shifts towards warming. On the example of the studied event, it was also shown that the use of different variants of the definition results in a different severity of an event, expressed via its intensity and geographical extent.

Naturally, the choice of temperature metric used to identify extreme events depends on the goal of the research being performed, and the comparison discussed herein is designed to serve the purpose of the evaluation of differences in results for reference only. The differences between the indices based on TX, TN, and their intersection are in the average number of hot days, strength and statistical significance of trends, and their spatial variability. The only fixed relationship is the fact that the number of hot days, as determined with the use of the TX and TN intersection, is always about one-half of the number of hot days identified with the use of TX or TN separately. This is consistent with results obtained by Lavaysse et al. (2018) who showed that the number of hot days identified using the 90th percentile of TN and of TX in Europe in the years 1995–2015 was almost the same, whereas the number of hot days identified with the use of an intersection of TN and TX was less than half as many. It would be rather futile to look for rules in the remaining cases, because for indices based on TX and TN, the average number of hot days is either higher or lower and the strength, and even the trend direction, is different depending on the region of Europe that is

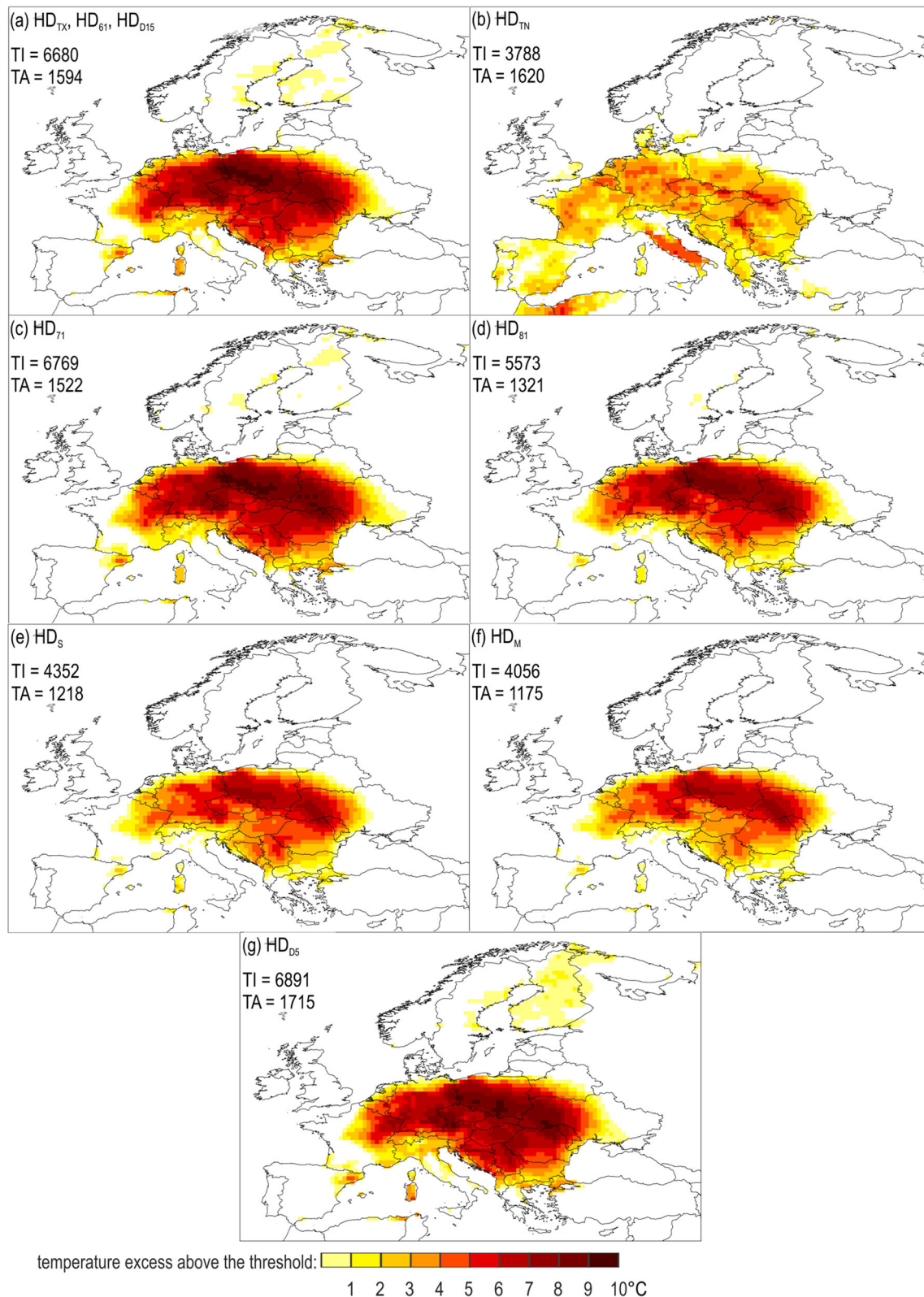


Fig. 8 Spatial extent (total area, TA) and intensity (total intensity, TI) of the extreme temperature event of August 31, 2015, shown with the use of different research approaches. Abbreviations explained in Table 1

investigated. This explains why Fenner et al. (2018) discovered that positive trends in heat waves in Berlin were twice as

high for an index based on TN in comparison with that based on TX and, on the other hand, Croitoru et al. (2016) found that

an increase in the frequency of heat waves in Romania was more rapid when they were defined using TX instead of TN.

If indices based on different baseline periods, namely 1961–1990, 1971–2000, and 1981–2010, are considered, it has been shown that usually, although not in every region of Europe, the use of an older baseline period results in a higher average number of hot days and a higher rate of change in their frequency. While rather similar results are obtained for indices based on two older periods, the index based on the most recent one is characterized by a relatively small average number of hot days, by different spatial distributions, and by a substantially slower increase rate in their frequency of occurrence. This is a direct effect of warming, as percentile-based thresholds determined using this period are relatively high and they rarely exceeded during the entire study period. On the example of the grid point “KRK” located near Kraków, Poland, it has been shown how general warming and temperature conditions during baseline periods affect percentile-based thresholds. Similar patterns were obtained by Croitoru et al. (2016) at some Romanian weather stations; likewise, at KRK, the latest considered baseline period was clearly warmer than the others, and the largest differences between thresholds reached 2.3 °C at KRK and 2.0 °C in Eastern and Southeastern Romania.

It has been shown that differences in results obtained with the use of percentiles calculated for different timeframes within the year are relatively small. The spatial variability of the average number of hot days and also geographical patterns of trends including their direction, strength, and statistical significance are highly comparable among the indices in this group. Differences become apparent in analyses of the distribution of the number of hot days within a season. As shown in the case study, this results in a different intensity and spatial extent of a temperature extreme event, depending on a chosen index and the summer day considered. The advantage of thresholds based on daily percentiles over seasonal and monthly ones is that they are relevant for any part of the season/year. As a result, they enable to identify temperature extremes throughout the whole year without sudden changes in the threshold values at the turn of seasons or months. A 5-day window is the basis for indices developed by the Expert Team on Climate Change Detection and Indices (ETCCDI) (for details, see Alexander et al. 2006); however, as has been shown, the threshold determined with the use of this window is characterized by relatively high day-to-day variability. This may be the simple reason why many researchers choose to use a window of larger size, for example, a 15-day window (for example, Della-Marta et al. 2007a; Fischer and Schär 2010; Perkins et al. 2012).

It has been repeatedly emphasized in climatological studies that the broadly accepted definition of a temperature extreme is very general and that there is a huge number of indices for the analysis of such events. The indices used are based on

different research approaches, and the decision to use a given variant of the definition is frequently related to impact groups or sectors considered. Nevertheless, sometimes, researchers are interested in an analysis of the climatology of extremes in itself, which means multiannual variability and trends. In each case, one should be aware of the implications of the use of a given research approach, including its potential and limitations, to avoid misinterpretation of research results.

Funding information This research was funded by the Jagiellonian University in Kraków (grant no. K/DSC/004776).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Rupa Kumar K, Revadekar J, Griffiths G, Vincent L, Stephenson DB, Burn J, Aguilar E, Brunet M, Taylor M, New M, Zhai P, Rusticucci M, Vazquez-Aguirre JL (2006) Global observed changes in daily climate extremes of temperature and precipitation. *J Geophys Res* 111:1–22. <https://doi.org/10.1029/2005JD006290>
- Busuioc A, Dobrinescu A, Birsan MV, Dumitrescu A, Orzan A (2015) Spatial and temporal variability of climate extremes in Romania and associated large-scale mechanisms. *Int J Climatol* 35:1278–1300. <https://doi.org/10.1002/joc.4054>
- Carril AF, Gualdi S, Cherchi A, Navarra A (2008) Heatwaves in Europe: areas of homogeneous variability and links with the regional to large-scale atmospheric and SSTs anomalies. *Clim Dyn* 30:77–98. <https://doi.org/10.1007/s00382-007-0274-5>
- Croitoru AE, Piticar A, Ciupertea AF, Roșca CF (2016) Changes in heat waves indices in Romania over the period 1961–2015. *Glob Planet Change* 146:109–121. <https://doi.org/10.1016/j.gloplacha.2016.08.016>
- Della-Marta PM, Haylock MR, Luterbacher J, Wanner H (2007a) Doubled length of western European summer heat waves since 1880. *J Geophys Res* 112:1–11. <https://doi.org/10.1029/2007JD008510>
- Della-Marta PM, Luterbacher J, von Weissenfluh H, Xoplaki E, Brunet M, Wanner H (2007b) Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcings and predictability. *Clim Dyn* 29:251–275. <https://doi.org/10.1007/s00382-007-0233-1>
- Efthymiadis D, Goodess CM, Jones PD (2011) Trends in Mediterranean gridded temperature extremes and large-scale circulation influences. *Nat Hazards Earth Syst Sci* 11:2199–2214. <https://doi.org/10.5194/nhess-11-2199-2011>
- Fenner D, Holtmann A, Krug A, Scherer D (2018) Heat waves in Berlin and Potsdam, Germany – Long-term trends and comparison of heat

- wave definitions from 1893 to 2017. *Int J Climatol* 1–16. <https://doi.org/10.1002/joc.5962>
- Fischer EM, Schär C (2010) Consistent geographical patterns of changes in high-impact European heatwaves. *Nat Geosci* 3:398–403. <https://doi.org/10.1038/ngeo866>
- Fu-Min REN, Trewin B, Brunet M, Dushmanta P, Walter A, Baddour O, Korber M (2018) A research progress review on regional extreme events. *Adv Clim Chang Res* 9:161–169. <https://doi.org/10.1016/j.accre.2018.08.001>
- Haylock MR, Hofstra N, Tank AK, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J Geophys Res* 113(D20119). <https://doi.org/10.1029/2008JD010201>
- Horton RM, Mankin JS, Lesk C, Coffel E, Raymond C (2016) A review of recent advances in research on extreme heat events. *Curr Clim Change Rep* 2:242–259. <https://doi.org/10.1007/s40641-016-0042-x>
- Hoy A, Hänsel S, Skalak P, Ustrnul Z, Bochníček O (2017) The extreme European summer of 2015 in a long-term perspective. *Int J Climatol* 37:943–962. <https://doi.org/10.1002/joc.4751>
- Intergovernmental Panel on Climate Change (2012) Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the intergovernmental panel on climate change [Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner G-K, Allen SK, Tignor M, Midgley PM (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA
- Kažys J, Stankunavičius G, Rimkus E, Bukantis A, Valiukas D (2011) Long-range alternation of extreme high day and night temperatures in Lithuania. *Baltica* 24:72–81
- Kysely J (2010) Recent severe heat waves in Central Europe: how to view them in a long-term prospect? *Int J Climatol* 30(1):89–109
- Lavaysse C, Cammalleri C, Dosio A, van der Schrier G, Toreti A, Vogt J (2018) Towards a monitoring system of temperature extremes in Europe. *Nat Hazards Earth Syst Sci* 18:91–104. <https://doi.org/10.5194/nhess-18-91-2018>
- Lhotka O, Kysely J (2015a) Hot Central-European summer of 2013 in a long-term context. *Int J Climatol* 35:4399–4407. <https://doi.org/10.1002/joc.4277>
- Lhotka O, Kysely J (2015b) Characterizing joint effects of spatial extent, temperature magnitude and duration of heat waves and cold spells over Central Europe. *Int J Climatol* 35:1232–1244. <https://doi.org/10.1002/joc.4050>
- Moberg A, Jones PD, Lister D, Walther A, Brunet M, Jacobeit J, Alexander LV, Della-Marta PM, Luterbacher J, Yiou P, Chen D, Klein Tank AMG, Saladié O, Sigró J, Aguilar E, Alexandersson H, Almarza C, Auer I, Barriendos M, Bergert M, Bergström H, Böhm R, Butler CJ, Caesar J, Drebs A, Founda D, Gerstengarbe F-W, Micela G, Maugeri M, Österle H, Pandzic K, Petrakis M, Srnc L, Tolasz R, Tuomenvirta H, Werner PC, Linderholm H, Philipp A, Wanner H, Xoplaki E (2006) Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000. *J Geophys Res* 111(D22106). <https://doi.org/10.1029/2006JD007103>
- Perkins SE (2015) A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmos Res* 164–165:242–267. <https://doi.org/10.1016/j.atmosres.2015.05.014>
- Perkins SE, Alexander LV (2013) On the measurement of heat waves. *J Climatol* 26:4500–4517. <https://doi.org/10.1175/JCLI-D-12-00383.1>
- Perkins SE, Alexander LV, Nairn JR (2012) Increasing frequency, intensity and duration of observed global heatwaves and warm spells. *Geophys Res Lett* 39:1–5. <https://doi.org/10.1029/2012GL053361>
- Revich B, Shaposhnikov D (2008) Excess mortality during heat waves and cold spells in Moscow, Russia. *Occup Environ Med* 65:691–696. <https://doi.org/10.1136/oem.2007.033944>
- Russo S, Sillmann J, Fischer EM (2015) Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environ Res Lett* 10:124003. <https://doi.org/10.1088/1748-9326/10/12/124003>
- Rusticucci M, Kysely J, Almeida G, Lhotka O (2016) Long-term variability of heat waves in Argentina and recurrence probability of the severe 2008 heat wave in Buenos Aires. *Theor Appl Climatol* 124: 679–689. <https://doi.org/10.1007/s00704-015-1445-7>
- Sanderson M, Economou T, Salmon K, Jones S (2017) Historical trends and variability in heat waves in the United Kingdom. *Atmosphere* 8: 191. <https://doi.org/10.3390/atmos8100191>
- Smith TT, Zaitchik BF, Gohlke JM (2013) Heat waves in the United States: definitions, patterns and trends. *Clim Change* 118:811–825. <https://doi.org/10.1007/s10584-012-0659-2>
- Spinoni J, Lakatos M, Szentimrey T, Bihari Z, Szalai S, Vogt J, Antofie T (2015) Heat and cold waves trends in the Carpathian Region from 1961 to 2010. *Int J Climatol* 35:4197–4209. <https://doi.org/10.1002/joc.4279>
- Stefanon M, Dandrea F, Drobinski P (2012) Heatwave classification over Europe and the Mediterranean region. *Environ Res Lett* 7. <https://doi.org/10.1088/1748-9326/7/1/014023>
- Stephenson DB (2008) Definition, diagnosis, and origin of extreme weather and climate events. In: Diaz HF, Murnane RJ (eds) *Climate extremes and society*. Cambridge University Press, New York, pp 11–23
- Sulikowska A, Walawender JP, Walawender E (2018) Temperature extremes in Alaska: temporal variability and circulation background. *Theor Appl Climatol*. <https://doi.org/10.1007/s00704-018-2528-z>
- Tomczyk AM, Bednorz E (2016) Heat waves in Central Europe and their circulation conditions. *Int J Climatol* 36:770–782. <https://doi.org/10.1002/joc.4381>
- Tomczyk AM, Pórolniczak M, Kolendowicz L (2018) Cold waves in Poznań (Poland) and thermal conditions in the city during selected cold waves. *Atmosphere* 9. <https://doi.org/10.3390/atmos9060208>
- Ustrnul Z, Wypych A, Kosowski M (2012) Extreme temperatures and precipitation in Poland - an evaluation attempt. *Meteorol Zeitschrift* 21:37–47
- von Storch H, Zwiers FW (2003) *Statistical analysis in climate research*. Cambridge University Press, Cambridge
- World Meteorological Organization (2009) *Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation*. [Tank AK, Zwiers FW, Zhang X (eds)]. Geneva, Switzerland
- World Meteorological Organization (2017) *WMO guidelines on the calculation of climate normals*. Document WMO-No. 1203. https://library.wmo.int/index.php?lvl=notice_display&id=20130#.XI-UIChKhaQ. Accessed 20 Sept 2019
- Wypych A, Sulikowska A, Ustrnul Z, Czekierda D (2017) Temporal variability of summer temperature extremes in Poland. *Atmosphere* 8. <https://doi.org/10.3390/atmos8030051>
- Xu Z, FitzGerald G, Guo Y, Jalaludin B, Tong S (2016) Impact of heatwave on mortality under different heatwave definitions: a systematic review and meta-analysis. *Environ Int* 89:193–203
- You Q, Jiang Z, Kong L, Wu Z, Bao Y, Kang S, Pepin N (2017) A comparison of heat wave climatologies and trends in China based on multiple definitions. *Clim Dyn* 48:3975–3989. <https://doi.org/10.1007/s00382-016-3315-0>
- Zhang X, Alexander L, Hegerl GC, Jones P, Klein Tank A, Peterson TC, Trewin P, Zwiers FW (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Clim Change*. <https://doi.org/10.1002/wcc.147>